基于策略改进遗传算法的双星星族光谱拟合[1]

迟焕斌[1]，李忠木[2]，王锋[1,3]

(1.昆明理工大学管理经济学院，云南昆明 650093; 2. 大理大学天文研究所，云南大理 671000; 3．广州大学天体物理中心，广东广州 510006)

摘 要：双星星族光谱拟合的实质是在双星星族理论光谱库对应的参数空间进行搜索，通过最小二乘法找到与被拟合观测光谱最相近的理论光谱。快速、准确的拟合是有效利用星族合成方法处理海量星系光谱的难点和关键。为了提高双星星族光谱拟合的速度，针对双星星族光谱拟合计算量巨大、拟合效率不高等问题，本文在双星星族光谱拟合的优化模型基础上，采用策略改进的遗传算法对模型进行求解，并与 BS2fit 算法和传统遗传算法进行比较。实验表明，通过改进策略的遗传算法对双星星族光谱拟合的速度平均可提高 43.5%，一定程度上推动了演化星族合成方法在天文研究中的应用。

关键词：光谱拟合；星系；遗传算法

中图分类号：P1-092; TP391　　文献标识码：A　　文章编号：

0 引 言

双星星族是包含一些共同演化的双星和单星，具有相同初始金属丰度和演化年龄的恒星组合，它是对单星星族的扩展和丰富。基于双星星族模型得到的星团和星系参数结果通常更合理[1]，因此，基于双星星族的研究得到越来越多的关注。

光谱拟合是获得星系物理信息的重要方法[2]。随着各种巡天项目的开展和天文信息技术的发展，人们获得了越来越多的高质量光谱数据，为天体光谱拟合创造了条件，有效地推动了光谱拟合的发展。光谱拟合近年来取得了一系列成果，如文献[2]成功地用贝叶斯方法对单星星族光谱拟合估计出星系参数信息。文献[3]基于贝叶斯证据用贝叶斯模型选择的方法分析了双星星族和单星星族模型，发现双星星族模型比其它没有考虑双星相互作用的模型的光谱拟合效果好很多。文献[4]和[5]对早期星系紫外到可见波段的光谱拟合发现双星星族模型对球状星团蓝离散星观测结果有很好的解释。文献[6]Han, Z. (2007)首次通过双星星族光谱拟合发现它能很好地解释早型星系的紫外反转现象。

由于双星星族演化比传统单星星族更复杂，考虑的物理输入也更多，其对应的理论光谱库很大，一定拟合精度要求下的光谱采样率也更高，所以双星星族光谱拟合需要进行海量的计算。能否快速高效地进行双星星族光谱拟合直接影响着星族演化合成方法的应用。因此，探索基于双星星族模型的快速有效的光谱拟合方法显得十分重要。

为了解决双星星族光谱拟合速度不能满足星系演化等研究需求的问题，加快星族演化合

＊　基金项目：国家重点研发计划(2018YFA0404603),国家自然科学基金会-中国科学院天文联合基金资助项目(U1831204，U1931141)，国家自然科学基金资助项目(1196114001，11903009，　11863002，　GZ1284) ,云南省重点研发计划（2018IA054, 云南省汪景琇院士工作站

收稿日期: 2020-06-15　　　　修订日期: 2020-07-13

作者简介:迟焕斌，　男，　博士研究生，　研究方向：数据处理及智能计算。Email:563372780@qq.com

通讯作者:王　锋，　男，　教授。研究方向：海量天文数据处理. Email: fengwang@gzhu.edu.cn

成方法在天文研究中的推广和运用，本文基于管理策略进行了一个尝试。我们将光谱拟合转化为一类最优化问题，以最小化理论光谱和观测光谱之间的均方误差为优化目标，建立最优化模型，提出改进策略的遗传算法对模型求解，并与 BS2fit 算法[1]和传统遗传算法对比以验证算法的有效性。

1 双星星族模型介绍

双星星族模型的一个主要目标是构建高精度且完备的双星星族理论光谱库。双星星族的理论光谱构建包括两个部分：模型输入参数的假设和理论光谱的计算。

如文献[1]所述，双星星族模型采用 Hurley[7]的快速恒星演化程序代替恒星演化轨迹，采用 BaseL3.1 恒星光谱库[8]和 Chapbrier 初始质量函数[9]，并假设每个双星的主星和次星质量比 q 服从 0－1 的均匀分布[9]。恒星形成率假设为$\psi(t)=\psi(0)\cdot e^{-t/\tau}$[8]的形式，$t$和$\tau$分别表示星系的年龄和恒星形成的持续时标。

利用演化星族合成方法，在初始质量函数和恒星形成率已知的情况下可以计算星系中不同时刻形成的恒星数。然后，借助 Hurley 的快速恒星演化程序来进行恒星演化计算。最后通过 BaSeL3.1 恒星样本库将恒星的演化参数转换为光谱能量分布。在完成恒星演化计算后将所包含的单颗恒星的光谱进行合成，就可以计算年龄为t的星族在波长λ的光谱流量$f_{\lambda0}$。每个星族光谱流量由三个参数决定:金属丰度 Z，年老星族年龄t_1，和年轻星族年龄t_2。考虑恒星速度弥散σ_*的影响，尘埃消光（采用 Calzetti 等人在 2000 年提出的消光理论[8]），红移，银河系消光等因素，双星星族合成中在波长λ处观测的光谱流量可表达为：

$$f_\lambda=[f_{\lambda0}(z,t_1,t_2)\otimes G(v_*,\sigma_*)]p(\tau_\lambda)r(A_\lambda)\#(1)$$

其中，$f_{\lambda0}$ 为由三个参数(Z, t_1, t_2)决定星族的光谱流量，G(v_*,σ_*)是平均值和标准偏差分别为v_*和σ_*一个高斯分布，p(τ_λ)是通过尘埃消光为τ_λ的尘埃的能量百分比，r(A_λ)是通过银河系消光A_λ后的能量比，⊗为卷积算子。

为了简化双星星族的理论光谱库，本文将v_*,σ_*，τ_λ,A_λ的取值均设为 0，仅考虑 Z，t_1，t_2。然而这并不影响实验目的，同样可以对比分析出所提出的该进遗传算法在光谱拟合中的表现，因为我们所采取的对比实验也是基于同样的简化的双星星族理论光谱库。

2 基于策略改进遗传算法的双星星族光谱拟合

2.1 双星星族光谱拟合模型

光谱拟合的目的是为了获得星系星团的物理参数信息。如上一节所述，在双星星族模型中光谱是由z,t_1,t_2，v_*,σ_*，τ_λ,A_λ 和星族恒星质量 8 个参数共同决定的。其中，星系的恒星质量可由 Thomas 等(2005)[10]的拟合公式计算:

$$\log\left(M_*/M_\odot\right)=0.63+4.52\log\left(\sigma_0/\left(\text{kms}^{-1}\right)\right)\tag{2}$$

式中的M_*为星系的恒星质量，σ_0是恒星的速度弥散。其它参数可通过光谱拟合，即将星族的理论光谱与观测光谱进行比对从而给出星族组成的定量信息。具体任务是从理论光谱库中找到最接近观察光谱的理论光谱。在光谱拟合中，通过将观测光谱与理论光谱进行比对，并最小化理论光谱与观测光谱数据之间的差异以确定拟合最优的理论光谱，进而推断出恒星组成的主要参数。

双星星族理论光谱库有稠密的参数网格：恒星金属丰度(Z)取值为 0.0003~0.03，当 Z < 0.001 时其参数间隔为 0.0001，金属丰度较高时(Z > 0.001)参数间隔设为 0.001。星族群的年老星族年龄(t_1)在 0 ~ 15Gyr 范围内，其间隔设为 0.1Gyr；年轻星族年龄(t_2)从 t_1 递减到 0Gyr，其间隔设为 0.1Gyr。恒星速度弥散 σ_* 参数以步长为 10 km s^{-1} 从 0 到 350km s^{-1} 改变；尘埃消光(τ_λ)取值是时间隔为 0.01 的 0 至 0.5 区间。因此，双星星族理论光谱库包含巨大参数空间，光谱拟合包含天然的海量计算，对此建立拟合模型如下。

记观察光谱为$f_{\mathrm{ob}\lambda}$ ，双星星族合成理论光谱为$f_{\mathrm{th}\lambda}$，则光谱拟合契合度的可由

$$\chi^2 = \Sigma[(f_{\mathrm{ob}\lambda} - f_{\mathrm{th}\lambda})^2\omega_\lambda] \#(3)$$

刻画。其中ω_λ是波长 λ 的权重参数。最佳拟合结果对应于 min(χ^2)。在大多数情况下，ω_λ使用观测误差$E_{\mathrm{o}\lambda}^2$表示：

$$\omega_\lambda = 1/E_{\mathrm{o}\lambda}^2 \# (4)$$

ω_λ也可以根据特殊要求赋予某个阈值。星族光谱拟合的实质是在理论光谱库对应的参数空间进行搜索，找到与被拟合的观察光谱χ^2最小值所对应的理论光谱。当χ^2最小时，双星星族合成光谱拟合模型满足物理约束：0<ω_i<1，其本质是一个最优化问题，可形式化为：

$$\begin{cases} \min & \chi^2 \\ \text{s.t.} & w_i > 0 \\ & \sum w_i = 1 \end{cases}$$

其中，(3)式为目标函数。

2.2 双星星族光谱拟合算法

遗传算法[11]（GA，Genetic Algorithm）是模拟自然界生物遗传进化行为，提出选择、交叉和变异算子，通过对求解问题空间的多点并行迭代搜索能快速有效的收敛至全局最优解的一类仿生启发式算法。遗传算法是一种基于种群的搜索算法，其个体代表了可能解的一个样本。随着搜索的迭代进行，个体通过分享和交换他们之间的信息来产生更好的后代。传统的遗传算法（CGA, Conventional Genetic Algorithm）包含选择、交叉和突变等至关重要的操作。他们可以选择获胜的个体，维持种群的多样性，并在种群内的个体之间交换信息。

然而，传统改进策略遗传算法在局部搜索能力上表现较差，且对初值非常敏感。对 GA 算法运行机制分析后，受粒子群算法惯性自适应的启发，同时为克服该不足，本文在传统遗传算法改进策略的基础上设计了一种自适应遗传算法（SGA, Self-adaptive Genetic Algorithm），来求解双星星族光谱拟合问题。传统遗传算法（CGA）采用固定交叉概率（PC, Probability of Cross）和固定突变概率（PM, Probability of Mutation）。这使得算法按照既定的规则进行，从而导致遗传变异不够灵活。基于特定的进化遗传使得算法的性能如收敛速度和全局优化能力对进化参数的选择很敏感。如果个体的适应度值大于总体的平均适应度值，则分配的交叉概率和变异概率应该较小。此外，对适应度值越大的个体，表示该个体越接近目标解，因此，应赋予其较小的交叉和变异概率，以保护其信息不被遗传给下一代。与之相反，对适应值较小的个体则应伺机引入扰动因子增强其变异能力，个体应被赋予较大的交叉和突变概率，以增加其变异概率。于是，适应度值越小的个体变异的机会越大。为了减少进化参数选择的依赖性，提高算法的收敛速度和全局优化能力，SGA 改进如下：

1) 自适应 PC 更新：

$$\begin{cases} p_{\mathrm{c}} = \begin{cases} C_2\left(1-\frac{(1-\gamma_{\mathrm{CA}})^{w}(f_c-f_{\mathrm{ovs}})}{f_{\max}-f_{\mathrm{avs}}}\right); & f_{\mathrm{c}} > f_{\mathrm{avg}} \\ C_2; \quad f_{\mathrm{c}} \leqslant f_{\mathrm{avg}} \end{cases} \\ \qquad w = \left(\frac{2}{\mathrm{iteration}}\right)^{0.3} \end{cases} \tag{5}$$

2） 自适应 PM 更新：

$$p_{\mathrm{m}} = \begin{cases} M_2\left(1-\frac{(1-\gamma_{\mathrm{GA}})^{n}(f_{\max}-f_{\mathrm{m}})}{f_{\max}-f_{\mathrm{avg}}}\right); & f_{\mathrm{m}} > f_{\mathrm{avg}} \\ M_2; \quad f_{\mathrm{m}} \leqslant f_{\mathrm{avg}} \end{cases} \tag{6}$$

其中，$iteration$为当前迭代次数，f_{avg}为种群的平均适应值，f_c为选择交叉操作的个体适应值，f_{m}为选择变异操作的个体适应值，fmax为最大适应值，γ_{GA}为随机参数，据经验取$\gamma_{\mathrm{GA}}=0.5$。

3） 干扰因子主要目的是增强算法对搜索空间的探索能力。为此，引入了如下扰动因子（DO, Disruption Operator）：

$$D_{op} = \begin{cases} Dis_{i,j} \times \delta(-2,2) & \text{if} \quad \mathrm{Dis}_{i,\,\mathrm{best}} \geq 1 \\ 1 + Dis_{i,\,\mathrm{best}} \times \delta\left(\frac{-10^{-4}}{2}, \frac{10^{-4}}{2}\right) & \text{otherwise} \end{cases} \tag{7}$$

其中：δ是(-2,2)之间的随机数， $\mathrm{Dis}_{i,j}$是第 i 个染色体和领域第 j 个染色体的汉明距离。$Dis_{i,\,\mathrm{best}}$是当前最优染色体和第 i 个染色体的汉明距离。

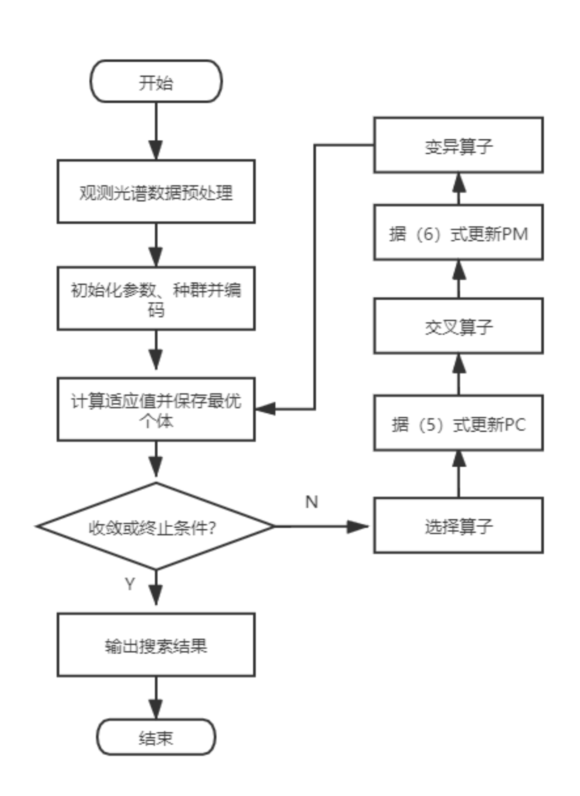


图 1 SGA 流程图

Fig. 1 SGA flow chart

SGA 算法具体流程如图 1 所示，主要包括以下步骤：

步骤一：光谱数据预处理,包括剔除发射线、归一化和红移处理；

步骤二：问题编码与种群初始化；

步骤三：按目标函数计算适应值并保留当前最优种群。决定种群的大小包括染色体长度和个数等，依策略初始化种群，设定进化终止条件；

步骤四：判断是否达到收敛条件或停止条件，如果满足则输出搜索结果，否则进行迭代进化；

步骤五：选择算子:根据适应度值随机选择一组较好的个体；

步骤六：按（5）式更新交叉概率；

步骤七：交叉算子:从集合中随机选择两染色体交换一个或多个位，根据交叉概率产生两个新的个体，按（7）式添加扰动；

步骤八：按（6）式更新变异算子：从集合中随机选择一个染色体，并根据突变概率改变一个或多个位以产生一个新的染色体；

步骤九：返回步骤四进行终止判断，循环以上步骤，直到终止条件。

3 实验与分析

3.1 改进策略的遗传算法实验

为了验证SGA在双星星族光谱拟合中的寻优效率，本文选取了CGA和文献[1]的BS2fit算法作为对比。实验代码采用Python3.6开发，操作系统为ubuntu18.04 64bit，内存容量为8 GB，CPU 型号为 i7-4790,主频3.60GHz。实验的理论光谱数据使用基于Chapbrier初始函数的双星星族理论光谱库进行拟合[1]。观测光谱数据选自99个本地星系的紫外-可见波段的光谱数据，可以在互联网上下载[2]。

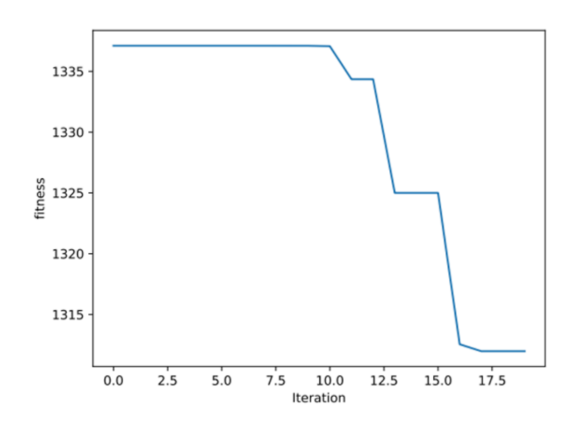
这些数据的特点是覆盖从紫外到可见光波段的波长，范围很广。此外，由于双星对紫外光谱有明显的影响，这些数据也是探索双星演化对光谱拟合影响的理想数据。每条观察光谱采样点数是N_λ=3871，以5500Å为归一化因子，对光谱进行预处理。SGA算法参数设置为：交叉概率上限C2=1,变异概率上限M2=1，随机参数$\gamma_{GA}=0.5$，个体编码长度n=20，种群规模Pop=100。耗时指标取每个算法重复20次的平均值。

实验首先从观测光谱数据中随机选取一条(NGC1399星系)作为实验数据参考文献 [12] 进行数据预处理；然后将观察光谱进行坏点、发射线、天光线等屏蔽得到编号为spec-0266-51630-0010的光谱数据；接着，借鉴相关研究[13]，采用χ^2/N_λ作为评价标准， 当$\chi^2/N_\lambda<1$时, 认为理论光谱与观测光谱拟合良好， 星族合成参数估计准确，拟合结果可以接受。拟合迭代过程如图2左所示。算法在种群迭代20次后收敛，对应的适应值χ^2=1308，拟合精确度为χ^2/N_λ=0.3433，理论光谱与观测光谱拟合良好，如图2右所示。其中灰色曲线代表实际观察的光谱，红色曲线代表所拟合得的最优理论光谱。

---

1. http://cluster.shao.ac.cn/~zhongmu/ssp_chab_hrseds/

2. ftp://ftp.stsci.edu/pub/catalogs/nearby_gal/sed.html



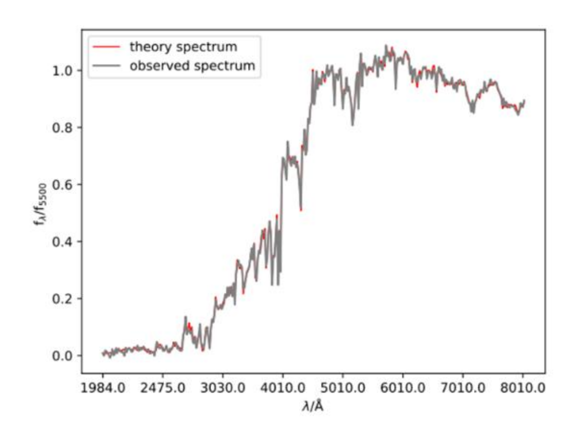
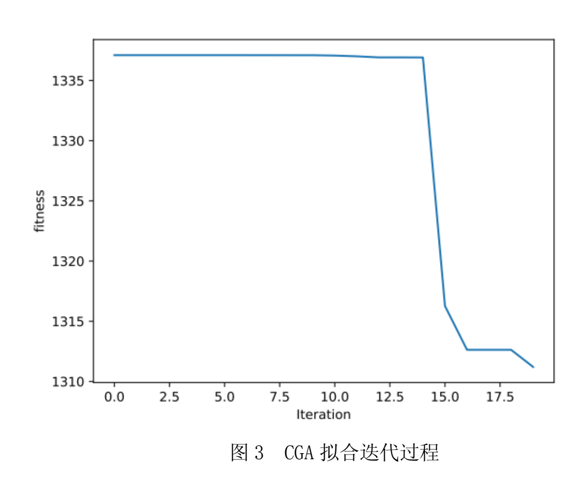


图 2　SGA 拟合迭代过程

Fig.2 SGA fit

3.2 与传统遗传算法（CGA）的比较



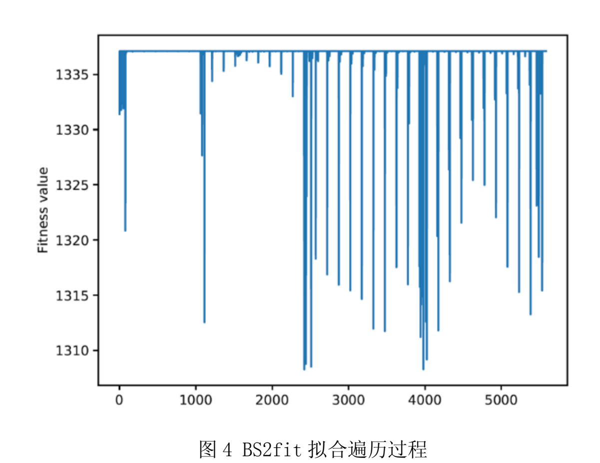


图 3　CGA 拟合迭代过程

图 4 BS2fit 拟合遍历过程

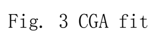

Fig. 3 CGA fit

Fig.4 BS2fit fit

CGA 算法参数：交叉概率 PC=0.8，变异概率 PM=0.1，染色体长度为 20，Pop=100。对测试光谱用 CGA 算法在双星星族理论光谱库上进行拟合结果如图 3 所示。CGA 算法中，种群迭代 16 次后陷入局部最优，迭代(iteration)15 次后开始收敛，寻到拟合度最优的理论光谱。实验迭代 20 次平均耗时 3201.3s 比 BS2fit 算法略为占优，速度提高了 19.2%。与 CGA 算法相比，SGA 在迭代 20 次后收敛，能有效跳出局部最优陷阱，收敛速度显著改善，实验 20 次平均耗时为 1304.4s，速度提高了 19.1%。

3.3 与 BS2fit 算法的比较

BS2fit算法采用文献[1]所述的BS2fit算法python实现，拟合遍历过程如图4所示，横轴表示遍历计算的次数，纵轴表示每次计算得到的卡方统计量的数值。由于是全网格搜索，双星星族模型的理论库覆盖波段长、演化参数网格稠密（模型光谱库覆盖金属丰度Z从0.0003到0.03，精度为0.0001；t1、t2范围从0到15Gyr，精度为0.1 Gyr）等导致样本参数空间巨大。虽然我们将模型参数做了简化处理，但是要完成一条观测光谱的拟合也需要逐一进行海量对比，能寻到拟合度最优的理论光谱。实验20次平均耗时为3954.2s。与BS2fit算法相比，SGA

算法具有更强的全局搜索能力，对理论光谱搜索效率更高，收敛速度显著改善，20次平均耗时减少了67.2%。

如上，三种算法均能找到最小χ^2对应的理论光谱，按照理论光谱的参数网格设置，反演得到表 1 中给出的拟合参数。

表 1：NGC1399 星系 bsSP-fitted 拟合参数

Tab.1 bsSP-fitted Parameters of NGC1399 Galaxies

星系参数	Z(金属丰度)	t_1年老星族年龄	t_2年轻星族年龄
本文拟合值	0.0250	6.30	6.30
文献[4] bsSP-fitted 拟合值	0.030± 0.00	13.50 ± 0.00	12.27 ± 0.81

分析产生误差的原因是我们对实验模型进行了简化，一是理论光谱库只考虑（Z，t_1，t_2）而没有考虑v_*,σ_*，τ_λ,A_λ等参数的影响；二是没有考虑观测光谱的误差，没有对观测光谱的不确定性进行处理，即是信噪比（S/N）取值很大的理想情况。文中得到的金属丰度比文献[4]的拟合值小 0.005。这也符合我们的预期，原因是模型输入的不确定性通常会导致金属丰度估计值偏低[4]。星族年龄（t_1，t_2）的拟合值比文献[4]的值低 50%的主要原因是简化的模型没有考虑速度弥散从而导致星族年龄被低估。该拟合结果也体现了“年龄-金属丰度简并效应”。

综上，增加自适应策略的改进遗传算法在拟合速度上较传统遗传算法和BS2fit算法平均提高了43.5%，拟合结果符合预期。

4 总结与讨论

采用自适应策略改进的遗传算法对双星星族光谱拟合，结果表明：（1）采用群智能算法（SIA）解决双星光谱拟合问题是可行的。（2）改进策略遗传算法和传统遗传算法、BS2fit算法相比在双星星族光谱拟合中能将计算速度分别提升 19.1%和 67.2%，对双星星族光谱拟合速度平均可提高 43.5%，提升效果明显。改进策略的遗传算法在传统遗传算法的基础上，发挥遗传算法全局搜索能力强、相较于传统遗传算法收敛速度快。此外,改进策略的遗传算法在搜索时更智能,能有效跳出局部最优陷阱，提高遗传算法全局搜索能力,避免算法早熟,相对于 BS2fit 算法的全网格搜索 SGA 算法对搜索空间做了取舍所以效率更高。虽然改进策略遗传算法在双星星族拟合上提升了速度，但是还有很大提升空间。一方面，编码方式对算法探索能力和收敛性也有影响，改进算法编码如利用量子编码等也值得探索，另一方面，群智能算法发展日新月异，探索更有效的搜索策略，如探索和强化学习 Q-learning 等融合也很有必要。考虑到遗传算法易于并行实现，下一步可探索并行策略下的改进策略遗传算法对光谱拟合速度的改善。

参考文献:

[1] Li Z, Zhang L, Liu J, et al. Integrated spectral energy distributions of binary star composite stellar populations[J]. Monthly Notices of the Royal Astronomical Society, 2012, 424(2): 874-883.

[2] Han Y, Han Z. BayeSED: A GENERAL APPROACH TO FITTING THE SPECTRAL ENERGY DISTRIBUTION OF GALAXIES[J]. Astrophysical Journal Supplement Series, 2014, 215(1).

[3] Han Y , Han Z . A Comprehensive Bayesian Discrimination of the Simple Stellar Population Model, Star Formation History, and Dust Attenuation Law in the Spectral Energy Distribution Modeling of Galaxies[J]. The Astrophysical

Journal Supplement Series, 2018, 240(1).
[4] Li, Zhongmu , et al. Potential Importance of Binary Evolution in UV-Optical Spectral Fitting of Early-Type Galaxies." Astrophysical Journal 776.1(2013):37.
[5] Zhang F , Li L , Han Z , et al. Yunnan-III models for Evolutionary population synthesis[J]. Monthly Notices of the Royal Astronomical Society(4):3390-3408.
[6] Han Z, Podsiadlowski P, Lynasgray A E, et al. A binary model for the UV - upturn of elliptical galaxies[J]. Monthly Notices of the Royal Astronomical Society, 2007, 380(3): 1098-1118.
[7] Hurley J R, Tout C A, Pols O R, et al. Evolution of binary stars and the effect of tides on binary populations[J]. Monthly Notices of the Royal Astronomical Society, 2002, 329(4): 897-928.
[8] Calzetti D , Armus L , Bohlin R C , et al. The Dust Content and Opacity of Actively Star-Forming Galaxies[J]. Astrophysical Journal, 1999, 533(2):682.
[9] Chabrier G. The Galactic Disk Mass Function: Reconciliation of the Hubble Space Telescope and Nearby Determinations[J]. The Astrophysical Journal, 2003, 586(2).
[10] Thomas D, Maraston C, Bender R, et al. The Epochs of Early-Type Galaxy Formation as a Function of Environment[J]. The Astrophysical Journal, 2005, 621(2): 673-694.
[11] David E. Goldberg. Genetic Algorithm in Search, Optimization and Machine Learning[J]. addison wesley, 1989, xiii(7):2104–2116.
[12] 张茜, 张健楠, 赵永恒. 基于聚类的星系光谱分析. 天文研究与技术, 2020, 17(2): 233-243.
Zhang Xi, Zhang Jiannan, Zhao Yongheng. Spectral Classification of Galaxies Based on Clustering Analysis. Astronomical Research and Technology, 2020, 17(2): 233-243.
[13] 韩金姝.基于遗传算法的星族合成参数估计[J].天文学报,2015,56(02):93-101.
Han Jinshu. Parameter Estimation of Stellar Population Synthesis Using Genetic Algorithm [J]. Astronomy Letters, 2015,56(02):93-101.

Binary-Star Spectral Fitting Based on Strategy Improved Genetic Algorithm

Chi Huanbin[1], Li Zhongmu[2] , Wang Feng[1,3],

(1 . School of Management and Economics , Kunming University of Science and Technology , Kunming 650500 , China;2 . Dali University, Institute of Astronomy, Dali 671000, Yunnan ;3 .Center for Astrophysics ,Guangzhou University ,Guangzhou 510006 , China)

Abstract: The essence of the binary-star spectral fitting is to search the parameter space corresponding to the comparison of observed and theoretical spectra, via techniques such as least square. Rapid and accurate fitting of spectra is very important for studying a huge number of galaxies. Because it is not easy to deal with such problems as spectral fitting contains huge

amounts of calculation and the efficiency of fitting is not high. The strategic improved genetic algorithm is used for binary-star spectral fitting by this paper, and the comparison to BS2fit algorithm and traditional genetic algorithm is also shown. The experimental results show that the improved genetic algorithm can increase the spectral fitting speed by 43.5% on average, which can possibly promote the application of stellar population synthesis method in astronomical researches obviously.

Key words: Spectral fitting; Galaxies; Improved Genetic Algorithm